doi:10.1111/j.1365-2958.2008.06583.x

Genome-wide regulon and crystal structure of BlaI (Rv1846c) from *Mycobacterium tuberculosis*

Claudia Sala,[1] Ahmed Haouz,[2] Frederick A. Saul,[2] Isabelle Miras,[2] Ida Rosenkrands,[3] Pedro M. Alzari[2] and Stewart T. Cole[1]*

[1]*Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.*
[2]*Institut Pasteur, Département de Biologie Structurale et Chimie (URA 2185 CNRS), 75724 Paris, France.*
[3]*Department of Infectious Disease Immunology, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark.*

Summary

Comparative genomics with *Staphylococcus aureus* suggested the existence of a regulatory system governing beta-lactamase (BlaC) production in *Mycobacterium tuberculosis*. The crystal structure of Rv1846c, a winged helix regulator of previously unknown function, was solved thus revealing strong similarity to the BlaI and MecI repressors of *S. aureus*, which both respond to beta-lactam treatment. Using chromatin immunoprecipitation and hybridization to microarrays (ChIP-on-chip), the Rv1846c regulon was shown to comprise five separate genomic loci. Two of these mediate responses and resistance to beta-lactam antibiotics (*rv1845c*, *rv1846c–rv1847*; *blaC–sigC*); two encode membrane proteins of unknown function (*rv1456c*, *rv3921c*) while the last codes for ATP synthase (*rv1303–atpBEFHAGDC–rv1312*). The ChIP-on-chip findings were confirmed independently using electrophoretic mobility shift assays, DNAse footprinting and transcript analysis leading to Rv1846c being renamed BlaI. When cells were treated with beta-lactams, BlaI was released from its operator sites causing derepression of the regulon and upregulation of ATP synthase transcription. The existence of a potential regulatory loop between cell wall integrity and ATP production was previously unknown.

Accepted 15 December, 2008. *For correspondence. E-mail stewart.cole@epfl.ch; Tel. (+41) 21 693 1851; Fax (+41) 21 693 1790.

Introduction

With the advent of extensively drug-resistant strains of *Mycobacterium tuberculosis* (Gandhi *et al*., 2006) that are insensitive to both front-line and second-line therapy, it is imperative to find new drugs and drug combinations to treat tuberculosis patients (Anonymous, 2006). Beta-lactam antibiotics are arguably the most successful class of antimicrobial agents ever developed for use in humans but they have been of little effect against mycobacterial infections. There are several likely reasons for this and these include the highly impermeable cell envelope acting in conjunction with a beta-lactamase to restrict drug uptake and destroy the limited quantity of beta-lactam that enters the cell, respectively (Jarlier and Nikaido, 1990; Jarlier *et al*., 1991). When beta-lactam antibiotics were used together with the beta-lactamase inhibitors clavulanic acid or sulbactam, increased susceptibility of *M. tuberculosis* was observed. Likewise, some improvement was seen when imipenem, a beta-lactamase-stable drug, was employed (Chambers *et al*., 1995). However, while imipenem showed some efficacy against *M. tuberculosis* in both mice and humans, resistance was also encountered (Chambers *et al*., 2005).

Tubercle bacilli produce a class A beta-lactamase, which hydrolyses a range of beta-lactam antibiotics, and has been characterized extensively by biochemical and structural means (Wang *et al*., 2006; Hugonnet and Blanchard, 2007). In some Gram-positive bacteria, beta-lactamase production is regulated at the level of gene-expression by proteins that sense the presence of the antibiotic. This system was well described in *Bacillus licheniformis* (Zhu *et al*., 1990; Zhu *et al*., 1992) but is, perhaps, best exemplified in *Staphylococcus aureus* where different, yet similarly organized, regulatory systems govern the expression of the *blaZ*, beta-lactamase gene, or the methicillin-resistant penicillin-binding protein 2a gene, *mecA*. Linked to *blaZ* but expressed divergently from a bicistronic operon are *blaI*, encoding a winged helix repressor protein (Gregory *et al*., 1997), and *blaR1*, whose product is a transmembrane (TM) protein with two distinct domains. Likewise, adjacent to *mecA* are *mecI* and *mecR1*, which play similar roles to *blaI* and *blaR1* respectively. In the absence of inducer, the homodimeric BlaI/MecI proteins bind to related

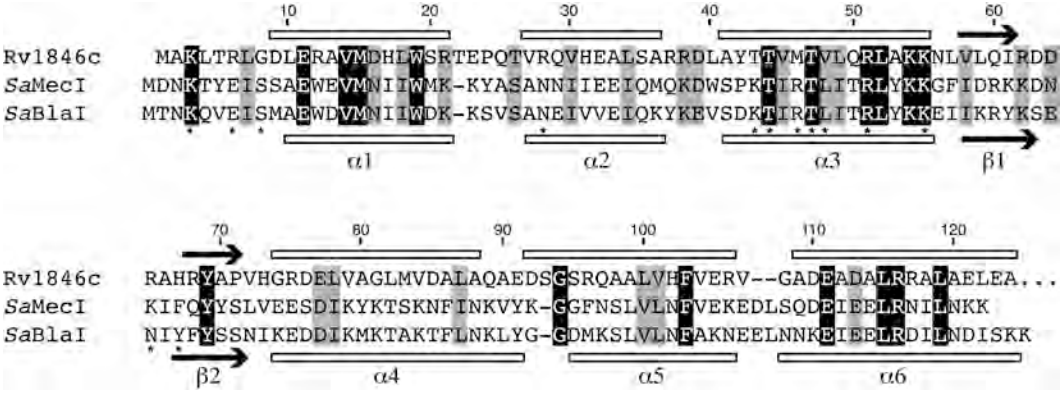

Fig. 1. Structure-based sequence alignment of Rv1846c (from Fig. 2) and *S. aureus* MecI and BlaI. Secondary structures of Rv1846c and BlaI are shown above and below the alignment respectively. Invariant amino acids are shown in black background, and positions with highly conservative substitutions in grey background. Each monomer consists of a N-terminal DNA-binding domain (helices α1-α3 and strands β1-β2) and a C-terminal dimerization domain (helices α4-α6). Residues involved in DNA binding are labelled with an asterisk below the sequence.

palindromic sites in the *blaZ/mecA* promoter regions to prevent their expression (Gregory *et al*., 1997; Garcia-Castellanos *et al*., 2004). The BlaR1/MecR1 proteins have similar organizations with four membrane-spanning segments and at least two biologically active domains. The extracytoplasmic domain binds beta-lactam molecules (Wilke *et al*., 2004) resulting in activation of the intracytoplasmic domain, a prometalloprotease that first cleaves BlaR1/MecR1 itself before proteolysing the cognate repressor (Zhang *et al*., 2001).

On inspection of the genome of *M. tuberculosis* (Cole *et al*., 1998), we detected two genes, *rv1846c–rv1845c*, which showed weak but convincing similarity to the respective *blaI–blaR1* and *mecI–mecR1* couplets. Although not linked to *blaC*, these were candidates for a potential beta-lactam-sensing signal transduction pathway, which is well conserved in mycobacteria, including *Mycobacterium leprae* whose genome has undergone massive gene decay (Cole *et al*., 2001). Consequently, we have purified the Rv1846c protein and solved its three-dimensional (3D) structure. Antibodies were raised against Rv1846c and used in ChIP-on-chip experiments to immunoprecipitate chromatin for hybridization to microarrays in order to identify the target genes (Grainger *et al*., 2005). The Rv1846c regulon was deduced from the ChIP-on-chip results and found to comprise at least five genomic loci, of which one (*sigC–blaC*) is directly involved in beta-lactam resistance while all respond to induction by the antibiotic.

Results

Rationale

The *Mycobacterium tuberculosis* gene *rv1846c* codes for a potential transcriptional regulator of 138 amino acids, which shares 23% and 21% identity with the staphylococcal repressors BlaI and MecI respectively (Fig. 1). The similarity of Rv1846c to the BlaI/MecI repressors suggested that it might play a role in beta-lactam resistance in *M. tuberculosis* and this possibility was supported by resemblance between the products of the downstream gene, *rv1845c*, and BlaR1 or MecR1, as discussed below. Furthermore, database searches revealed Rv1846c orthologues in all members of the *M. tuberculosis* complex, as well as in *Mycobacterium marinum*, *Mycobacterium ulcerans* and *Mycobacterium leprae* (Cole *et al*., 2001; Stinear *et al*., 2007; Stinear *et al*., 2008). Notably, the presence of the orthologue (ML2063) in the latter species suggests that the corresponding protein may play an important role, as it escaped the reductive evolution undergone by the leprosy bacillus (Cole *et al*., 2001). These intriguing similarities prompted us to further investigate the biological function of Rv1846c by means of biomolecular and structural analyses.

The crystal structure

Rv 1846c was crystallized as an asymmetric dimer (two crystallographically independent monomers) in space group $P2_12_12_1$ and as a symmetric dimer (one independent monomer) in space group $I2_12_12_1$. The structure of the asymmetric homodimer was determined using single-wavelength anomalous diffraction (SAD) techniques and refined to 1.8 Å resolution, while the second crystal form was solved by molecular replacement methods and refined to 2.7 Å (Table 1). As expected, the protein is a homodimer (Fig. 2a), whose structure is very similar to those observed for the repressors BlaI and MecI (Garcia-Castellanos *et al*., 2003; Safo *et al*., 2005). Each monomer consists of an N-terminal DNA-binding domain (DBD) that displays a characteristic winged-helix architecture (composed of helices α1-α3 and beta-hairpin β1-β2)

Table 1. Data collection, phasing and refinement statistics.

Data set	SeMet-labelled	Native
Data collection		
Data resolution (Å)[a]	30–1.8 (1.9–1.8)	50–2.7 (2.85–2.7)
Wavelength (Å)	0.9792	0.976
Space group	$P2_12_12_1$	$I2_12_12_1$
Cell dimensions (Å)	$a = 60.63$	$a = 51.06$
	$b = 64.36$	$b = 74.89$
	$c = 75.86$	$c = 75.89$
Unique reflections	27947	4243
Multiplicity[a]	3.5 (3.5)	6.9 (7.3)
Completeness (%)[a]	99.5 (99.4)	97.8 (99.3)
R_{sym} (%)[ab]	0.067 (0.493)	0.059 (0.510)
$\langle I/\sigma \rangle$[a]	11.7 (2.3)	10.3 (1.8)
Refinement		
Resolution (Å)	1.8	2.7
R_{cryst} [No. refs][c]	0.207 [25491]	0.219 [4048]
R_{free} [No. refs][c]	0.256 [1378]	0.282 [197]
Rms bonds (Å)	0.017	0.019
Rms angles (degrees)	1.57	1.76
Protein atoms	1876	910
Water molecules	244	–

a. Values in parentheses apply to the high resolution shell.

b. $R_{sym} = \sum_{hkl}\sum_{i}|I_i(hkl) - \langle I(hkl)\rangle| \Big/ \sum_{hkl}\sum_{i} I_i(hkl)$.

c. $R = \sum_{hkl}|F(h)\text{obs} - F(h)\text{calc}| \Big/ \sum_{hkl}|F(h)\text{obs}|$. R_{cryst} and R_{free} calculated from the working and test reflection sets respectively.

and a C-terminal dimerization domain composed of helices α4-α6 (Fig. 2a).

The DBD encompasses a central HTH motif (defined by helices α2 and α3) that binds to the major-groove. The domain terminates with a highly exposed β-hairpin (or β-wing) that in most cases also participates in protein–DNA interactions. This compact domain is stabilized by an extended hydrophobic core involving many residues that

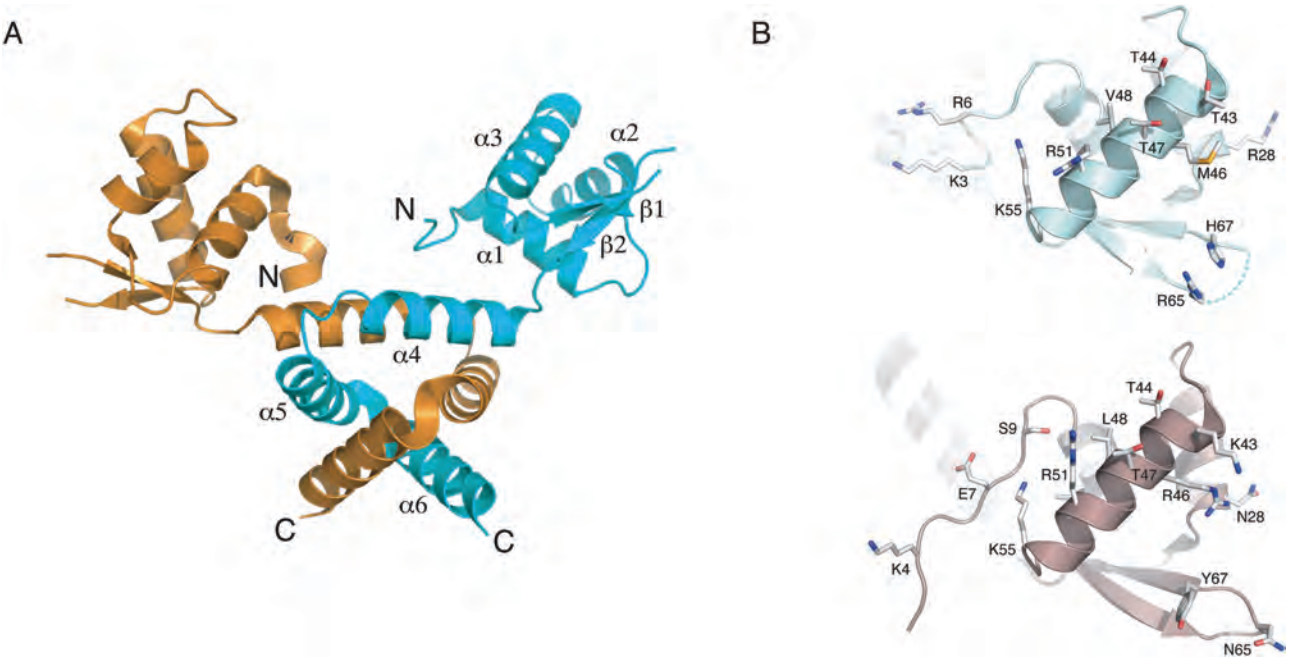

Fig. 2. Overall structure of Rv1846c.
A. Ribbon model representation of the Rv1846c homodimer in the orthorhombic crystal form. Secondary structure elements are labelled for one monomer.
B. A single DNA-binding subunit of Rv1846 (top) and BlaI (bottom), seen perpendicularly to the recognition helix (α3). DNA-contacting residues (as defined in Fig. 1) are shown in stick representation. Figure drawn with PyMol (DeLano, 2002).

Fig. 3. Conformational flexibility in Rv1846c. The DNA-binding winged-HTH domains (bottom) from the three crystallographically independent Rv1846c monomers (space group $P2_12_12_1$: cyan and violet, $I2_12_12_1$: yellow) are superimposed to that of *S. aureus* BlaI (green) bound to the DNA operator (pdb code 1XSD). Note the different orientations of the C-terminal dimerization domain.

are largely conserved in mycobacterial orthologues as well as in the BlaI/MecI family of repressors (Fig. 1). Core residues strictly conserved in mycobacteria include Leu7, Val14, Met15, Leu18, Trp19, Val30, Leu34, Leu40, Val45, Leu49, Leu52, Leu57, Val58 and Tyr69. The overall structure of the DBD is conserved in the three crystallographically independent copies of Rv1846c (rms deviations of Cα positions for residues 10–73 in the range 0.36–0.42 Å) and also resembles the equivalent domains of BlaI and MecI, with rms deviations of 1–1.2 Å (Fig. 3).

Despite the low sequence identity between the Rv1846c and BlaI/MecI DBDs (14–18%), the majority of residues important for DNA recognition in the staphylococcal repressors are largely conserved in the mycobacterial orthologue (Figs 1 and 2b). Thus, several DNA-contacting residues (namely Thr44, Thr47, Leu48, Arg51 and Lys55) in the recognition helix α3 from BlaI/MecI, which binds to the major groove of DNA and makes most specific DNA contacts (Garcia-Castellanos *et al.*, 2004; Safo *et al.*, 2005), are strictly conserved in Rv1846c, except for the substitution Leu48-Val (Fig. 1). Within the α3 helix, only BlaI Arg46 (which makes hydrogen-bonding contacts with the DNA phosphate backbone) is substituted by a strictly conserved methionine in mycobacterial Rv1846c orthologues. Also conserved outside the recognition helix is Lys3 (Lys4 in BlaI/MecI), which was found to be important for BlaI repression (Gregory *et al.*, 1997). Three other residues that were observed to interact with the DNA backbone in the BlaI/MecI-operator complexes (Glu7 at the N-terminus of the protein, Asn28 in helix α2 and Asn65 in the β-hairpin; BlaI numbering) are all substituted by a conserved arginine in Rv1846c orthologues (Fig. 1), suggesting that they could fulfill a similar phosphate-binding role.

The dimerization domain (DD) of Rv1846c, formed by helices α4-α6, displays the typical 'spiral staircase' architecture observed for the staphylococcal repressors (Fig. 2), although with some differences. In particular, the overall conformation of the Rv1846c monomer differs in the three crystallographically independent molecules, as shown in Fig. 3 upon superposition of the DBDs, and these differences are even greater when compared with BlaI in its DNA-bound conformation (Garcia-Castellanos *et al.*, 2004). This comparison reveals a conformational flexibility of the Rv1846c DBDs with respect to the DD, which may be important for DNA binding. Indeed, a similar conformational flexibility and hinge-bending movement have been observed for the unbound and DNA-bound forms of BlaI (Safo *et al.*, 2005) as well as for the (structurally similar) bacterial transcriptional regulators MexR (Lim *et al.*, 2002) and MtaN (Godsey *et al.*, 2001). The structure-based sequence alignment of Rv1846c, BlaI and MecI also shows local differences at the linker between helices α4 and α5 (Fig. 1). While all structures have a short linker segment, in Rv1846c helix α4 is one-turn shorter and helix α5 has an additional N-terminal turn. Also, the C-terminal region of Rv1846c is 12 or 15 residues longer than those of BlaI and MecI respectively. However, all residues beyond the last residue of helix α6 (Ala124, which coincides with the end of α6 in BlaI, Fig. 1) are disordered (and possibly flexible) in the Rv1846c crystal structures.

Defining the biological role of Rv1846c

The 3D structure indicates that Rv1846c is a transcriptional regulator and, as such, is expected to control expression of a certain number of genes. In order to define the Rv1846c regulon and unravel its biological function we performed chromatin immunoprecipitation and hybridization to microarrays (ChIP-on-chip). After hybridization, arrays were scanned and the signals plotted against their relative position on the *M. tuberculosis* H37Rv chromosome using Artemis (Rutherford *et al.*, 2000). Figure 4 and Table 2 present the results obtained from four independent experiments. A total of five putative Rv1846c binding sites were found throughout the genome. The first one (Table 2) was located in the intergenic region between *rv1846c* and *rv1847*, encoding a hypothetical protein with a

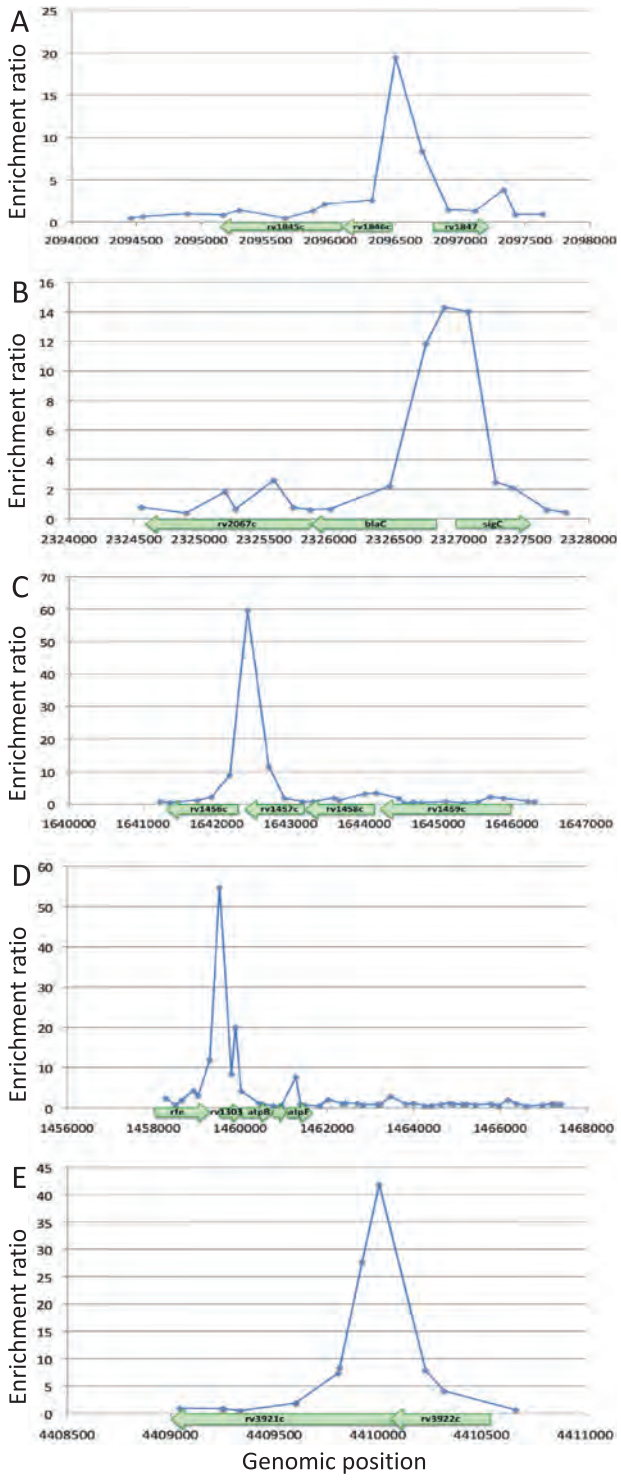

Fig. 4. Representative ChIP-on-chip results. DNA obtained from immunoprecipitation with anti-Rv1846c antibodies was labelled with Cy3 and hybridized to microarrays with Cy5-labelled genomic DNA. Data representing the five identified regulatory regions are plotted as Cy3/Cy5 ratios (*y*-axis) and as a function of the genomic position (*x*-axis). (A) *rv1845c–rv1847* locus; (B) *rv2067c–sigC* locus; (C) *rv1456c–rv1459c*; (D) *rfe–atpF* locus; (E) *rv3921c–rv3922c* locus.

thioesterase domain. As Rv1846c-like proteins usually control their own expression (Gregory *et al.*, 1997), this signal suggests that Rv1846c regulates its own transcription. Another binding site was noted in the intergenic region between *blaC*, which codes for the sole class A beta-lactamase in *M. tuberculosis*, and *sigC*, encoding an ECF subfamily sigma factor. Moreover, strong signals (Fig. 4; Table 2) were found upstream of *rv1456c*, which codes for an essential antibiotic-transporting ABC-transporter, and *rv1303*, encoding a membrane protein. Interestingly, *rv1303* partially overlaps the first gene of the ATP synthase gene cluster, *atpB*, thereby suggesting that they may all constitute a single transcriptional unit. Indeed, RT-PCR experiments demonstrated the existence of a polycistronic RNA covering both genes (Fig. S1). Thus, binding of Rv1846c upstream of *rv1303* contributes to the transcriptional regulation of the ATP synthase genes. Finally, a putative binding site was discovered upstream of *rv3921c*, predicted to encode another essential conserved TM protein. Taken together, these data suggest that Rv1846c controls a small subset of genes involved in regulation, antibiotic transport and detoxification, membrane components and, most surprisingly, genes responsible for ATP synthesis.

Confirming ChIP-on-chip results

In order to validate ChIP-on-chip results and confirm the ability of Rv1846c to bind to the corresponding DNA regions, electrophoretic mobility shift assays (EMSAs) were performed using specific DNA fragments covering the *rv1845c–rv1847* region and purified Rv1846c (see Fig 5A and B). Fragment 1, which encompasses the *rv1846c–rv1847* intergenic region, was clearly shifted, while fragment 3, which extends into the *rv1846c* coding sequence, bound Rv1846c to a lesser extent. By contrast, Rv1846c did not bind to fragments 2, 4 and 5, covering intragenic regions. These data confirmed ChIP-on-chip findings indicating the presence of a binding site in the region upstream of the *rv1846c* gene. Moreover, the protein–DNA interaction was found to be dose-dependent, as shown in Fig. 5C, suggesting the presence of more than one binding site or cooperative binding to DNA. Rv1846c was also able to bind to the orthologous *M. leprae* DNA region (coordinates 2 450 333–2 450 701), suggesting the existence of a conserved regulatory mechanism (Fig. 5C). Further EMSA experiments were performed to validate Rv1846c binding to the intergenic region between *blaC* and *sigC* (coordinates 2 326 777–2 326 965), and to the region upstream of *rv1456c* (coordinates 1 642 366–1 642 545). As shown in Fig. 5D, both DNA fragments were shifted in the presence of increasing amounts of the protein, thus confirming ChIP-on-chip results.

Table 2. Rv1846c regulon.

Gene	Function[a]	ChIP-on-chip score[b]
rv1846c–rv1847	Transcriptional regulator with penicillinase repressor domain and hypothetical protein belonging to the thioesterase family.	13.6
blaC–sigC	Class A beta-lactamase and RNA polymerase sigma factor (ECF subfamily).	13.5
rv1456c	Antibiotic-transport membrane ABC transporter.	49.2
rv1303–atpBEFHAGDC–rv1312	Membrane protein, ATP synthase and hypothetical secreted protein.	39.5
rv3921c	Transmembrane protein.	32.0

a. Gene function described according to TubercuList (http://genolist.pasteur.fr/TubercuList/).
b. Peak height was calculated as the ratio of immunoprecipitated DNA to genomic DNA. The mean values of four independent experiments are reported.

Genome-wide mapping of Rv1846c binding sites

Footprinting analysis was then performed in order to identify the Rv1846c operator site. The 355 bp BamHI–EcoRI DNA fragment was digested and labelled at the EcoRI site and incubated with increasing amounts of Rv1846c, as described in the legend to Fig. 5E. A clearly protected region, extending from position –32 to –63, was identified.

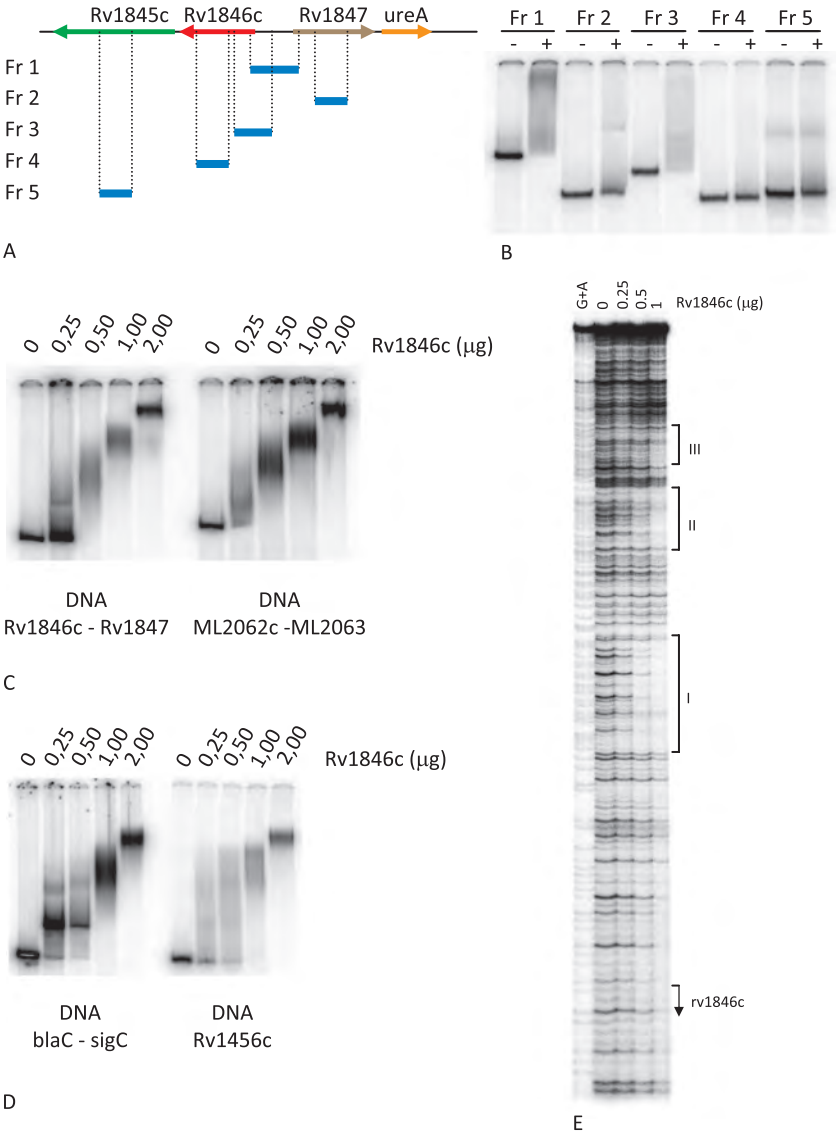

Fig. 5. Binding of Rv1846c to DNA.
A and B. (A) Schematic and (B) EMSA results showing binding of Rv1846c to the *rv1845c–rv1847* DNA locus. DNA fragments were incubated with 0.5 µg Rv1846c and run on 5% polyacrylamide gel.
C. Binding of Rv1846c to the *rv1846c–rv1847* and *M. leprae ML2062c–ML2063* intergenic regions.
D. Binding of Rv1846c to the *blaC–sigC* intergenic and *rv1456c* upstream regions.
E. DNAse footprinting of Rv1846c to the *rv1846c–rv1847* regulatory sequence. DNA was incubated with the indicated amount of protein and then treated with DNAse as described in the *Experimental procedures*. Maxam-Gilbert G+A sequence reaction was run alongside in a 6% denaturing gel. Protected regions are indicated by square brackets and the *rv1846c* start codon by a bent arrow.

This binding site is located upstream of *rv1846c*, likely in the region encompassing the 5′-regulatory sequences, thus confirming that Rv1846c controls its own expression. Moreover, at higher protein concentrations (0.5–1 μg) two more distal protected sequences were identified in the intergenic region between *rv1846c* and *rv1847*: the first extending from coordinate –102 to –128, and the second from –145 to –177. These findings were confirmed when the complementary strand was probed (data not shown).

The sequence of the protected region extending from coordinate –32 to –63 was used to define the Rv1846c operator sequence by multiple alignments of the cognate regions from *M. tuberculosis*, *M. leprae*, *M. marinum*, *M. ulcerans* and *M. smegmatis*.

In agreement with the protein structure, the operator site was found to be an inverted repeat with the following consensus: TACG/TAC<5>GTA/CGTA. This sequence was then used to scan the *M. tuberculosis* genome, thereby identifying all potential members of the regulon. Interestingly, the operator site was found in the appropriate position upstream of all of the target genes identified by ChIP-on-chip experiments, with the number of mismatches ranging from none (as in the case of *rv1303*) to two, as for *rv3921c* (Table 3). Other sites were identified by bioinformatics (Table S1) but, on re-examination of the ChIP-on-chip data, the corresponding genes did not appear to be part of the Rv1846c regulon, except for *cydA*, encoding a component of the respiratory chain, the integral membrane cytochrome D ubiquinol oxidase. This binding site was detected in three out of four ChIP-on-chip experiments but the calculated score was substantially lower than those of the other target genes.

Mapping 5′ ends of rv1846c–rv1847 *and* rv1303–atp *transcripts*

Further insight into the regulation of *rv1846c–rv1847* was achieved by mapping the 5′ ends of their transcripts using 5′ RACE on RNA extracted from *M. tuberculosis* H37Rv. A single PCR product was obtained for each gene and directly sequenced. The 5′ end of *rv1846c* was located 23 bases upstream of the translation start codon, thus giving a transcript containing a predicted ribosome binding site in its leader region, whereas the transcription start site of *rv1847* coincides with the GTG initiation codon, giving rise to a leaderless mRNA. The location of the potential promoters is consistent with the position of the previously identified Rv1846c operator site and the two regulatory regions likely overlap (Fig. S2).

Bioinformatic analysis predicted the binding site for Rv1846c to be located 241 bp upstream of the *rv1303* translation start codon. As this position is quite far from the gene, we investigated transcription of the *rv1303–atp* operon by RT-PCR and 5′ RACE. Results reported in

Table 3. Relationship between operator sequence and induction by beta-lactams analysed by ChIP-on-chip and quantitative RT-PCR.

Gene	Operator sequence[a]	Enrichment (IP DNA/GENOMIC DNA)[b]			Relative increment upon amoxicillin treatment (quantitative RT-PCR)[c]
		ChIP-on-chip score NT	ChIP-on-chip score Clavulanate 4 μg ml^{-1}	ChIP-on-chip score Amoxicillin 2 μg ml^{-1} + Clavulanate 4 μg ml^{-1}	
	5′-**TACGAC** – – – – **GTAGTA**-3′ **T** **C**				
rv1846c-rv1847	5′-**TACGAC**AGCGA**GTAGTA**-3′	3.2	3.2	1	*rv1846c*: 1.5 *rv1847*: 1.7
blaC–sigC	5′-**TACGAC**CGGTC**GTCGTA**-3′	5.5	5.5	3.5	*blaC*: 1.8 *sigC*: 1.4
rv1456c *rv1303–atpBEFHAGDC–rv1312*	5′-**TACTAC**AGCGT**GTAGTT**-3′ 5′-**TACGAC**AAGGA**GTAGTA**-3′	11 8	11 13	3 3	2.2 *rv1303*: 1.7 *atpB*: 1.9
rv3921c	5′-**TACGAC**CGGC**GTCGTG**-3′	9	9.5	2	3.7

a. Sequence motif identified in the *M. tuberculosis* H37Rv genome by the SEARCH PATTERN program in TubercuList.
b. Enrichment values represent the ratio between the immunoprecipitated DNA (IP) and total genomic DNA used as control.
c. Relative increment was calculated with the untreated sample as reference.
NT, not treated sample.
Bases in the operator are shown in bold.

Fig. S1 indicate the presence of a transcript with a long leader that extends far upstream of the *rv1303* start codon and whose 5′ end was mapped 194 bp upstream of the gene (Fig. S3). This is consistent with the location of the Rv1846c operator site and suggests an overlap with the promoter region. Bioinformatic analysis of the leader region ruled out the presence of un-annotated small open reading frames and the only predicted Shine-Dalgarno sequence was 8 bp upstream of *rv1303*, consistent with the position of the translational start codon.

Response to beta-lactam treatment

ChIP-on-chip was used to test whether Rv1846c, like BlaI and MecI, responds to beta-lactam antibiotics. The results from *M. tuberculosis* cells, mock-treated or treated with amoxicillin and clavulanate, or clavulanate alone, are presented in Fig. 6 and Table 3. Compared with the untreated sample, addition of clavulanate caused no decrease in signal, whereas amoxicillin treatment resulted in a marked reduction, indicating release of the repressor from the binding sites and presumably expression of the target genes.

These results were supported by quantitative RT-PCR analysis of RNA extracted from mock-treated or amoxicillin-treated *M. tuberculosis* cultures. Target transcripts were amplified in the presence of Sybr Green and relative expression calculated using *sigA* as housekeeping gene and the untreated sample as reference. Increased amounts of transcripts were detected for all of the target genes (Fig. 7), indicating the existence of a common regulatory mechanism. In particular, *rv1303* and *atpB*, which were shown to constitute a single operon, were both induced after treatment, confirming their coregulation. Moreover, there is an inverse correlation between the induction level observed upon antibiotic treatment and the conservation of the operator site: the more the gene is induced, the less the operator site conforms to the consensus (Table 3). Taken together, these data confirm that Rv1846c plays an important role in a regulatory system involved in response to cell wall stress and damage.

Discussion

In this work, we have established and validated ChIP-on-chip technology for *M. tuberculosis* by studying the genome-wide distribution of Rv1846c, a transcriptional regulator of the winged helix superfamily, whose 3D structure has also been determined in order to facilitate understanding of its function. ChIP-on-chip experiments identified the Rv1846c regulon as comprising five members, including *rv1846c* itself and genes involved in antibiotic transport, detoxification and cell wall function.

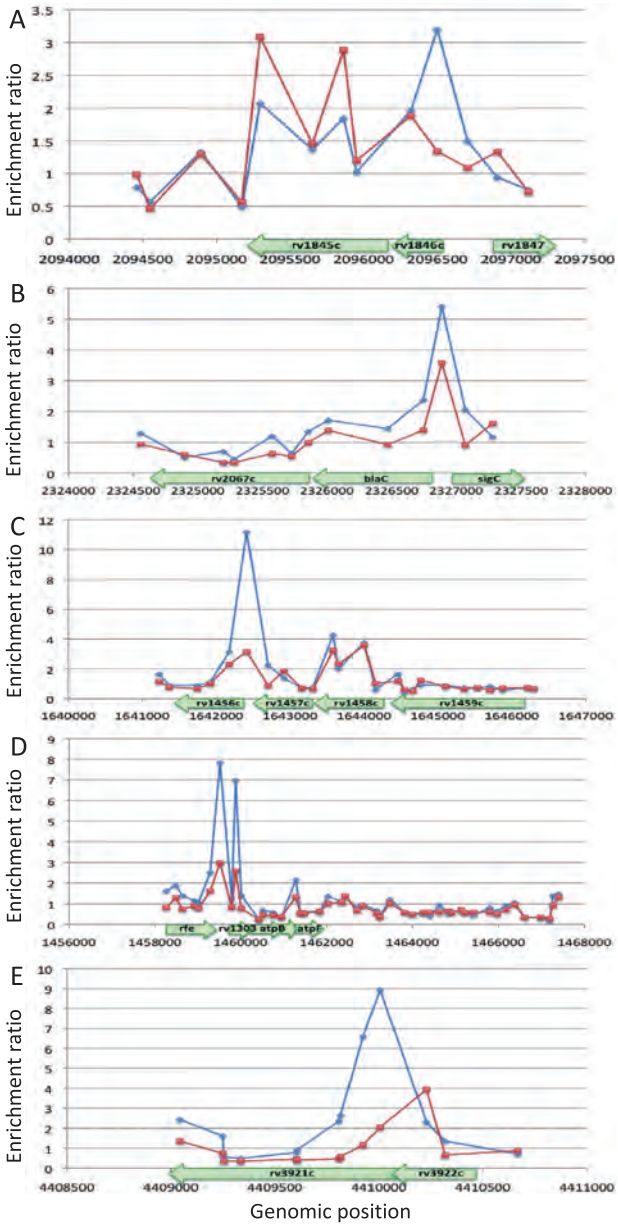

Fig. 6. ChIP-on-chip results after beta-lactam treatment. *M. tuberculosis* nucleoprotein extracts prepared from mock-treated or amoxicillin (2 μg ml^{-1}) plus clavulanate (4 μg ml^{-1})-treated cultures were used in immunoprecipitation experiments with anti-Rv1846c antibodies. Recovered DNA was labelled with Cy3 and hybridized to microarrays with Cy5-labelled genomic DNA. Data representing the five previously identified regulatory regions are plotted as Cy3/Cy5 ratios (*y*-axis) as a function of the genomic position (*x*-axis), with the blue line indicating the mock-treated sample and the red line the amoxicillin-treated. (A) *rv1845c–rv1847* locus; (B) *rv2067c–sigC* locus; (C) *rv1456c–rv1459c*; (D) *rfe–atpF* locus; (E) *rv3921c–rv3922c* locus.

Remarkably, one of the target sites is located upstream of the *rv1303–atpBEFHAGDC–rv1312* operon, arousing curiosity and interest about the regulation of the genes coding for ATP synthase. We will return to this point below.

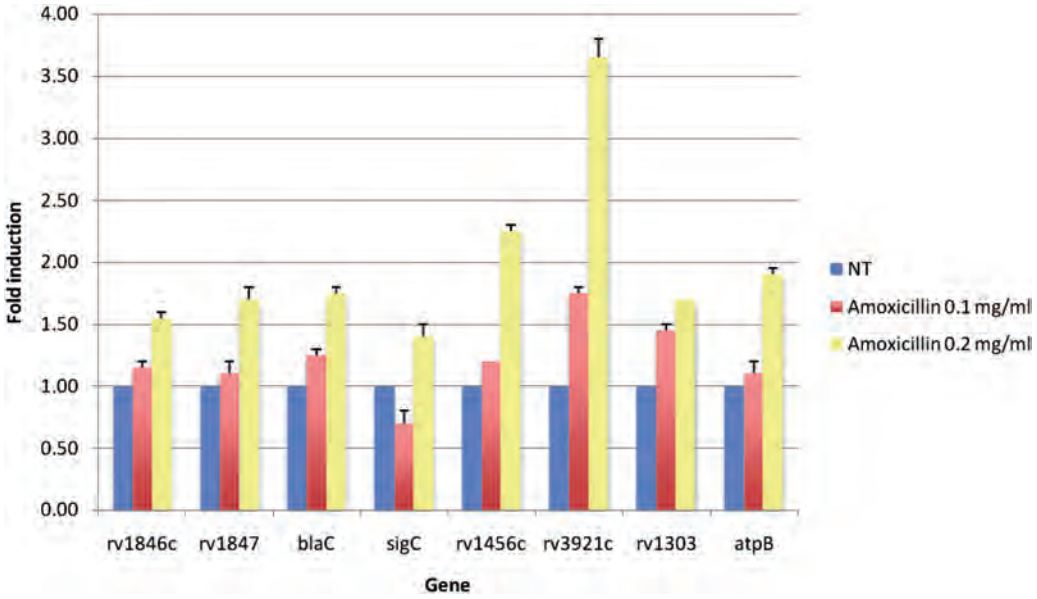

Fig. 7. Quantitative RT-PCR on RNA extracted from mock-treated or amoxicillin-treated *M. tuberculosis* cultures. Each gene was amplified with specific primers in the presence of Sybr Green and their relative expression was calculated using *sigA* as standard and the untreated sample as reference.

Immunoprecipitation results were confirmed independently by EMSAs and DNAse footprinting. On binding to its target regions, Rv1846c exhibited a typical pattern, characterized by multiple shifts at increasing protein concentrations, suggesting the presence of more than one binding site or a cooperative interaction with DNA. This observation was supported by footprinting experiments, where, in addition to a region proximal to the ORF, two more binding sites were identified between *rv1846c* and *rv1847*. These latter regions were protected at high protein concentrations and their sequences show very little conservation when compared with the sequence of the proximal site. In addition, no similar motifs were found upstream of the other Rv1846c-regulated genes, thus implying cooperativity and/or oligomerization of Rv1846c on binding to DNA, as previously reported for the *S. aureus* MecI repressor (Sharma *et al.*, 1998; Safo *et al.*, 2006). The site, extending from –32 to –63, clearly overlaps the *rv1846c* promoter region (Fig. S2) and Rv1846c thus controls its own expression. The aforementioned cooperativity displayed by the repressor on binding to DNA may also explain control of *rv1847*, with Rv1846c wrapping around DNA, as suggested for the MecI orthologue (Safo *et al.*, 2006).

Footprinting and multiple sequence alignments allowed the identification of the Rv1846c operator site as an inverted repeat, consistent with the protein acting as a dimer (Fig. 2), with each half-site resembling GTAGTA. Moreover, on screening the complete *M. tuberculosis* genome, the operator site was found upstream of all of the target genes identified by ChIP-on-chip. In particular, the sequence upstream of *rv1303* is the best conserved, with a perfect match, whereas the others have a variable number of mismatches, ranging from one (*rv1456c*) to two (*rv3921c*).

Re-examination of the ChIP-on-chip data sets after *in silico* analysis allowed the identification of an additional binding site corresponding to the *cydA* gene. The region upstream of *cydA* might represent another Rv1846c target but binding there could either be weaker or dependent upon specific culture conditions. Indeed, expression of *cydA* was induced upon beta-lactam treatment (up to 1.9-fold), thus representing another link between beta-lactam damage and respiration. The other genes, identified by bioinformatic tools but not by Chip-on-chip, may represent false positives. Alternatively, they might have escaped identification by immunoprecipitation because of inefficient cross-linking at some DNA loci or epitope masking due to particular nucleoprotein topologies. As expected, the operator sequence and Rv1846c regulon are conserved in all other sequenced mycobacterial genomes.

These observations were extended when the structural data on repressor-operator complexes were examined. Thus, the overall conservation of DNA-contacting residues in Rv1846c and BlaI/MecI (Figs 1 and 2b) strongly suggests a similar mode of DNA recognition. The analysis of the operator sequences recognized by Rv1846c and the conservation of these regions in mycobacterial genomes lend further support to this hypothesis. The conserved DNA sequence motif is palindromic, with its half-site (which binds a single HTH motif) having the sequence

GTAGTA, which is very similar to the reported BlaI/MecI recognition motif, GTAXT (Safo *et al*., 2005). The major difference appears to be the length of the spacer between the two half-sites, which is five base pairs in Rv1846c operators instead of four as in *blaI/mecI*. However, the observed flexibility of the DBDs in the Rv1846c homodimer (Fig. 3) could easily compensate for this single-base-pair shift, thus rendering possible a similar pattern of protein–DNA interactions as observed for the staphylococcal repressors. Taken together, the above observations strongly support a similar mode of action of staphylococcal BlaI/MecI and the mycobacterial Rv1846c-like repressors.

Extensive *in vivo* and *in vitro* studies have been performed to identify residues involved in protein–DNA interaction in Rv1846c orthologues. Wittman *et al*. (1993) demonstrated that deletion of six-amino-acid residues from the N-terminal region of the *B. licheniformis* penicillinase repressor eliminates the operator recognition and repression functions. In addition, site-specific mutagenesis in *S. aureus* BlaI and MecI proved the importance of the highly conserved residues Lys-4 and Asp-39, respectively, with the resultant mutant proteins being strongly compromised in their ability to bind to DNA and repress transcription (Gregory *et al*., 1997; Sharma *et al*., 1998).

All proteins from the BlaI/MecI family of bacterial repressors are inactivated by proteolytic cleavage at the highly conserved Asn-101–Phe-102 (BlaI numbering) peptide bond in the middle of helix α5 (Fig. 1). In Rv1846c orthologues, the sequence motif around the putative cleavage site is largely conserved except for the asparagine preceding the cleavage site, which is substituted by histidine in most mycobacterial sequences. This suggests a similar mode of repressor inactivation by proteolysis in mycobacteria, although the protease involved might have a different specificity.

Comparative genomics was used to identify a potential protease candidate in *M. tuberculosis*. The *S. aureus mec* and *bla* operons have a similar genomic organization, with the genes for the signal transducer and repressor forming a two-gene operon (*blaI–blaR1* and *mecI–mecR1*) that is divergently transcribed from its regulatory target gene (*blaZ* or *mecA*). The BlaR1/MecR1 proteins are proteases that cleave their cognate repressor following activation, and have a particular TM topology: three TM segments followed by a cytoplasmic protease domain, another TM segment and the extracellular beta-lactam-binding sensory domain. Interestingly, *rv1846c* is also part of a two-gene operon with *rv1845c*, and the latter's product has the same predicted topology as BlaR1/MecR1 (except for a missing C-terminal sensory domain) with which it also shares low overall sequence similarity. This resemblance, together with the presence of the peptidase family M48 motif (IPR001915), suggests that Rv1845c is likely to be a zinc-requiring metalloprotease, although its activation mechanism is different from those of BlaR1/MecR1. When *rv1845c* was interrupted by a transposon, the resultant mutant showed a slow growth phenotype (Sassetti and Rubin, 2003; Sassetti *et al*., 2003).

As Rv1845c (and all its mycobacterial orthologues) lacks an extracellular sensor domain, the antibiotic sensing mechanism operating in *M. tuberculosis* must be different: sensing may occur via interaction with the residual extracellular portion or could be mediated by other membrane proteins. Here, a possible candidate is the *M. tuberculosis* serine/threonine protein kinase, PknB, with its C-terminal PASTA domains, that have been implicated in interaction with beta-lactam antibiotics (Yeats *et al*., 2002; Wehenkel *et al*., 2008).

Direct evidence for Rv1846c responding to beta-lactam antibiotics was obtained using ChIP-on-chip and quantitative RT-PCR analysis. Upon amoxicillin treatment, Rv1846c detaches from its operator allowing transcription to take place. All of the genes in the regulon were de-repressed after antibiotic treatment and their induction level inversely correlated with the degree of conservation of the binding motif; the greater the conservation the weaker the induction observed. In addition, these results indicate that ChIP-on-chip technology can be exploited as a complementary approach to transcriptomics to define response to drug treatment or to different growth conditions. Based on our findings, we have renamed *rv1845c* and *rv1846c* as *blaR* and *blaI* respectively.

A further link to beta-lactams is the finding that BlaI controls expression of the *blaC* gene, encoding a class A beta-lactamase, which confers resistance to a variety of penicillins, cephalosporins and carbapenems (Hugonnet and Blanchard, 2007). Interestingly, *blaC* has been deleted from the *M. leprae* genome whereas the *blaR–blaI* orthologues (ML2064 and ML2063 respectively) are still present (Cole *et al*., 2001). This indicates that the beta-lactamase is not required by *M. leprae* and suggests that the signal transduction system might be involved in different physiological processes, like responding to peptidoglycan damage or to ATP synthase inhibitors. Co-expressed with *blaC* in *M. tuberculosis* is the *sigC* gene encoding an alternative RNA polymerase sigma factor, required for the transcription of a regulon important in pathogenesis (Sun *et al*., 2004; Karls *et al*., 2006), which includes *senX* and *mtrA*, members of two-component systems (Sun *et al*., 2004). Transcription of these two genes was not affected by beta-lactam treatment (data not shown).

The most intriguing and unexpected result of our work is the regulation of the *rv1303–atpBEFHAGDC–rv1312* operon by BlaI and the beta-lactam, amoxicillin. Transcript analysis (Fig. S1) showed *rv1303* to be cotranscribed with

atpB as part of a polycistronic RNA and orthologues of *rv1303* precede the *atp* genes in all sequenced mycobacterial genomes. The *rv1303* gene, which encodes a membrane protein of unknown function, is restricted to mycobacteria (Fig. S4) and appears to be essential for the growth of *M. tuberculosis* in macrophages (Rengarajan *et al*., 2005). Following *atpC* by 8 bp is *rv1312*, which codes for a 147 aa membrane protein, and this gene occurs in the *atp* operon of all sequenced actinobacteria (Fig. S4). It is curious that the gene arrangement for such an important and ubiquitous function as ATP synthase should be so atypical in mycobacteria.

Our findings have uncovered a possible link between ATP production and beta-lactam-induced cell wall damage, and BlaI/BlaR may act as a bridge between them. Peptidoglycan biosynthesis requires ATP and the cell wall damage arising from its inhibition might perturb the proton gradient across the membrane thereby reducing ATP synthase activity. Restoring ATP synthase function, in order to provide sufficient ATP for survival and, later on, for growth, would require *de novo* synthesis of the enzyme. Indeed, using transcriptomics, Boshoff *et al.* (2004) reported upregulation of *blaI* (*rv1846c*) when *M. tuberculosis* was treated with respiratory and ATP synthase inhibitors, thus suggesting that BlaI is involved in the response to different physiological stresses. In the absence of stress, a basal transcription level is guaranteed and translation of the polycistronic *rv1303–atpBEFHAGDC–rv1312* transcript would ensure production of ATP synthase and thus sufficient ATP levels inside the cell. Based on a model proposed previously for *S. aureus* (Wilke *et al*., 2004), BlaI might be cleaved by BlaR upon inhibition of ATP synthase and the *rv1303–atpBEFHAGDC–rv1312* operon would then be de-repressed, so that the circuit returns to steady state. It is particularly interesting that a *blaR* (*rv1845c*) mutant that fails to make the BlaR protease grows very slowly (Sassetti *et al*., 2003) and this may be explained by constant repression of *rv1303* by BlaI resulting in a shortage of ATP synthase. This hypothesis, and the effect of ATP synthase inhibitors, will be investigated further.

Experimental procedures

Bacterial strains and culture conditions

Mycobacterium tuberculosis H37Rv was grown in Dubos Broth (Difco) supplemented with Middlebrook albumin-dextrose-catalase (ADC) enrichment and 0.05% Tween 80 or on solid Middlebrook 7H11 medium (Difco) supplemented with oleic acid-albumin-dextrose-catalase (OADC). Beta-lactam treatment was performed by adding amoxicillin (Sigma) or amoxicillin plus clavulanate (Sigma) to exponentially growing cultures at the concentrations indicated in the *Results* section and incubating at 37°C for 3 h.

Oligonucleotides and plasmids

Plasmids and oligonucleotides used in this study are reported in Tables S2 and S3.

Chromatin immunoprecipitation experiments

Mycobacterium tuberculosis cultures (50 ml) grown up to OD_{600} 0.4–0.6 were treated with formaldehyde (final concentration 1%) and incubated 10 min at 37°C. Cross-linking was quenched by addition of glycine (final concentration 125 mM). Cells were then collected by centrifugation, washed twice with Tris-buffered saline (20 mM Tris-HCl pH 7.5, 150 mM NaCl) and stored at −80°C. Pellets were re-suspended in 4 ml Immunoprecipitation buffer (50 mM Hepes-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, Roche Antiprotease mini) and sonicated in a water bath sonicator (Bioruptor, Diagenode) to shear DNA to an average size of 300–700 bp. Cell debris was removed by centrifugation and the supernatant was used as input sample in immunoprecipitation experiments. Approximately 3.2 ml of the input sample was incubated overnight at 4°C on a rotating wheel with 20 µl of serum containing anti-Rv1846c rat polyclonal antibodies. An immunoprecipitation experiment without antibody was also set up as negative control. Protein–DNA complexes were immunoprecipitated with 80 µl of Dynabeads Sheep anti-Rat IgG (Dynal Biotech) for 90 min at 4°C. The magnetic beads were then collected and washed twice with immunoprecipitation buffer, once with immunoprecipitation buffer plus 500 mM NaCl, once with wash buffer III (10 mM Tris-HCl pH 8, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet-P40, 0.5% sodium deoxycholate), and once with Tris-EDTA buffer (pH 7.5). Immunoprecipitated complexes were eluted from the beads by treatment with 100 µl elution buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% SDS) at 65°C for 20 min. Samples were then treated with RNase A and cross-links were reversed by incubation for 2 h at 56°C and 6 h at 65°C in 0.5× elution buffer plus 50 µg proteinase K. DNA was extracted twice with phenol-chloroform, precipitated and re-suspended in 20 µl of water.

Microarray design, labelling and hybridization

Microarrays were designed and manufactured by Oxford Gene Technology (OGT), using Agilent technology. Arrays consist of 22 575 60-mer oligonucleotide probes, covering the genomes of the *M. tuberculosis* strains H37Rv and CDC1551, and the *M. bovis* strains AF2122/97 and BCG Pasteur. On average, there is an oligonucleotide probe every 203 bases. DNA obtained from immunoprecipitation and total genomic DNA were labelled by Klenow random priming, with Cy3-dCTP or Cy5-dCTP (GE Healthcare), respectively, and purified with Qiagen MinElute kit, according to the manufacturer's protocol. Labelled DNA was hybridized to microarrays in an Agilent Technologies Microarray chamber at 55°C for 72 h in the following buffer: 1 M NaCl, 50 mM MES, 20% formamide, 20 mM EDTA, 1% Triton X-100. The arrays were then washed once in 6× SSPE (0.18 M NaCl, 10 mM phosphate pH 7.4, 1 mM EDTA), 0.005% *N*-lauryl sarcosine and once in 0.06% SSPE, 0.18% polyethylene glycol 200, both for

5 min at room temperature. Finally, microarrays were dried in isopropanol and immediately scanned.

Microarray scanning and data analysis

For data acquisition, we used an Agilent Technologies microarray scanner and results were extracted by using Genepix Pro 5.0 software. The Cy3/Cy5 intensity ratio was calculated for each spot and plotted against the corresponding genomic position on the H37Rv chromosome in the Artemis software (Rutherford *et al*., 2000).

The physical arrangement of probes on the array does not reflect their location on the *M. tuberculosis* chromosome; data are put in the genomic order after scanning of the microarrays. Because DNA fragments obtained from ChIP experiments were between 300 and 700 bp, they were expected to hybridize to different probes adjacent to each other on H37Rv genome. A putative binding site was defined as a stretch of two or more consecutive spots with a > 2-fold Cy3/Cy5 ratio in all of the four independent experiments. Genes were described according to TubercuList (http://genolist.pasteur.fr/TubercuList/).

Electrophoretic mobility shift assay

DNA probes for gel shift experiments were amplified by PCR with specific primers containing either a BamHI or EcoRI site at the 5′ end, labelled with the Klenow enzyme and [α-^{32}P]-dATP, and purified with the Qiaquick Nucleotide Removal kit, according to the manufacturer's instructions (Qiagen). Binding reactions were performed in 10 µl of binding buffer (10 mM Tris-HCl pH 8, 50 mM KCl, 5% glycerol, 50 µg ml^{-1} BSA, 1 mM EDTA, 50 µg ml^{-1} of salmon sperm DNA) with about 5 fmol of the probe and different amounts of Rv1846c protein, for 20 min at room temperature. Reactions were loaded on a 5% polyacrylamide gel and run at 150 V for 2 h at room temperature. Gels were then dried and exposed to PhosphorImager (Molecular Dynamics) detection.

Footprinting experiments

DNA probes carrying BamHI and EcoRI restriction sites were obtained by PCR amplification, digested with either BamHI or EcoRI, labelled with the Klenow enzyme and [α-^{32}P]-dATP, and purified. Binding reactions were performed as described for the EMSA experiments, with about 60 000 c.p.m. of probe in a final volume of 50 µl. DNase I digestion was carried out by treatment with 1 µl of 0.12 mg ml^{-1} DNase I and 2 mM $CaCl_2$ for 1 min at room temperature. Reactions were stopped by addition of 25 µl of DNase stop buffer (0.1 M EDTA, 0.8% SDS, 1.6 M NH_4-Acetate, 300 mg ml^{-1} salmon sperm DNA), ethanol precipitated and re-suspended in 12 µl of sample buffer (7 M urea, 20 mM Tris-HCl pH 8) with tracking dyes. Samples were then loaded on a 7 M urea, 6% polyacrylamide sequencing gel. Maxam-Gilbert G+A sequencing reactions were performed as described previously (Maxam and Gilbert, 1977).

RNA extraction

Forty millilitres of *M. tuberculosis* H37Rv cultures (OD_{600} = 0.4) were pelleted and cells re-suspended in 100 µl Tris-EDTA buffer plus 75 µl RNA lysis buffer (4 M guanidinium thiocyanate, 0.01 M Tris pH 7.5, 0.97% β-mercaptoethanol). The suspension was sonicated in a water bath sonicator (Bioruptor, Diagenode) for 15 min (cycles of 30 s ON and 30 s OFF) at 4°C in 1.5 ml tubes. RNA was then purified with the SV Total RNA Isolation System, according to the manufacturer's protocol (Promega). Samples were subsequently extracted once with phenol-chloroform using Maxtract High Density Gel (Qiagen) and ethanol precipitated. After DNase treatment and a second phenol-chloroform extraction, RNA was re-suspended in water and stored at –80°C. The amount and purity of RNA was determined spectrophotometrically.

5′ RACE

Two micrograms of *M. tuberculosis* H37Rv RNA and 1 µg of primer 69, 72 or 120 (for *rv1846c*, *rv1847* and *rv1303* 5′ end mapping respectively) were incubated at 70°C for 5 min and then at 55°C for one hour in the presence of 1× cDNA synthesis buffer, 1 mM dNTPs, 40 U RNase inhibitor, 25 U Transcriptor Reverse Transcriptase (Roche). cDNA was then purified with the High Pure PCR Product Purification kit (Roche), and used in the subsequent poly(A) tailing reaction (30 min at 37°C in the presence of 0.2 mM dATP and 80 U Terminal Transferase, Roche). Semi-nested PCR amplification on poly(A)-tailed cDNA was performed using an oligo dT-anchor primer and primers 70, 73 or 109 for *rv1846c*, *rv1847* and *rv1303* respectively. Only one amplification product for each gene was obtained and directly sequenced.

RT-PCR

Two micrograms of *M. tuberculosis* RNA was incubated with 2.5 µM random primers at 70°C for 5 min. After cooling on ice, 1 mM dNTPs, 40 U RNase inhibitor and 25 U Transcriptor Reverse Transcriptase (Roche) were added in a final volume of 20 µl. The reaction was incubated at room temperature for 10 min and at 55°C for 30 min. cDNA was then used in the next PCR amplification, using oligonucleotides 67, 68, 77, 78, 107.

Quantitative RT-PCR

Two micrograms of RNA were reverse-transcribed with random primers and Transcriptor Reverse Transcriptase as described above. Sybr Green PCR Master Mix (Applied Biosystems) was used to quantify target cDNA molecules in a 20 µl reaction containing 100 nM of each primer, 10 µl Master Mix, 1 µl cDNA. Each reaction was run in duplicate in an Applied Biosystems 7900HT Sequence Detection System with the following protocol: denaturation at 95°C for 3 min, 40 cycles of denaturation at 95°C for 30 s and annealing/elongation with data collection at 60°C for 40 s. Melting curves were constructed to ensure that only one amplification product was obtained. Relative expression of target genes was determined according to the ΔΔCt method, with *sigA* as standard, using the following calculation: $\text{Ratio} = [(1 + \text{Efficiency of amplification}_{\text{sigA}})^{\text{Ct(treatment)}} : (1 + \text{Efficiency of amplification}_{\text{gene of interest}})^{\text{Ct(treatment)}}] : [(1 + \text{Efficiency of amplification}_{\text{sigA}})^{\text{Ct(no treatment sample)}} : (1 + \text{Efficiency of amplification}_{\text{gene of interest}})^{\text{Ct(no treatment sample)}}]$.

Recombinant protein production

The Rv1846c coding sequence was amplified by two-step PCR from the genomic DNA of *M. tuberculosis* H37Rv (Cole *et al.*, 1998) and cloned into the expression vector pDEST17 (Invitrogen). Transformed BL21(DE3)pLysS cells were used to inoculate 500 ml of LB medium containing 100 μg ml^{-1} ampicillin and 25 μg ml^{-1} chloramphenicol. The culture was grown for 4 h at 30°C and induced with 1 mM isopropopyl β-D-thiogalactoside (IPTG) at OD_{600} of 0.8–0.9. After 1.5 h of induced growth, cells were harvested by centrifugation, re-suspended and frozen. Cellular suspension was thawed and, after cell lysis and centrifugation, the supernatant was loaded on Ni-NTA resin (Qiagen) and the protein eluted by a linear gradient of imidazole. The purified protein (at 0.5 mg ml^{-1} in 50 mM Tris-HCl buffer with 0.1 M NaCl and 1 mM β-mercaptoethanol, pH 8.0) was incubated with His_6-tagged TEV protease at a protein-TEV ratio (w/w) of 1:7 for 6 h at 30°C and the reaction stopped by centrifugation. The supernatant was loaded on Ni-NTA resin and the flow-through (containing the tag-free protein) was loaded on a calibrated HiLoad 16/60 Superdex 75 gel filtration column (Amersham Biosciences) previously equilibrated with the dialysis buffer. Peak fractions were pooled and concentrated to 14.4 mg ml^{-1} (10 mg ml^{-1} for the SeMet-labelled protein) for crystallization. Dynamic light scattering and size exclusion chromatography suggest that Rv1846c is a dimer in solution, as observed for BlaI and MecI.

Generation of antiserum to Rv1846c

Four Lewis rats were immunised subcutaneously with 15 μg recombinant Rv1846c in Montanide ISA720 adjuvant (Seppic, Paris, France) in a total volume of 0.3 ml. On days 13 and 34, the rats received a booster immunisation, and at days 26 and 47, serum samples were collected. The serum samples were evaluated for recognition of recombinant Rv1846c by Western blot analysis, using alkaline phosphatase conjugated goat-anti-rat IgG (A8438, Sigma-Aldrich, Brøndby, Denmark) for detection, a serum-pool from naïve rats was used as a negative control. All serum samples collected from immunised rats showed a positive response, and were thereafter pooled and used as antiserum to Rv1846c.

Crystallization and structure determination

Initial robotic crystallization screenings were carried out at 18°C with a Cartesian Technology workstation, using sitting drops composed of 200 nl of native protein (10 mg ml^{-1}) and 200 nl of mother liquor equilibrated against 150 μl of the well solution on Greiner plates. Two different crystal forms of the protein were obtained at 18°C and manually optimized. The best diffracting crystals belong to space group $P2_12_12_1$ (with two monomers in the asymmetric unit) and were obtained in 1.6 M NaCl, 50 mM $MgCl_2$ and 100 mM Na citrate, pH 5.6. The second form (space group $I2_12_12_1$ with one monomer in the asymmetric unit) was grown in 2.4 M Na acetate and 100 mM Na cacodylate, pH 6.5. Diffraction data sets for the selenomethionine-labelled protein in space group $P2_12_12_1$ and for the native protein in space group $I2_12_12_1$ were collected using synchrotron radiation on beamline ID14-4 at the European Synchrotron Radiation Facility, Grenoble (France). For data collection, crystals were transferred to a cryoprotectant solution containing the mother liquor + 25% (v/v) of glycerol. Crystallographic parameters and data collection statistics are shown in Table 1.

The structure was determined using SAD data from a $P2_12_12_1$ crystal of SeMet-labelled Rv1846c, measured at the K edge of selenium. The substructure of the Se atoms was solved using program SOLVE (Terwilliger and Berendzen, 1999), locating six sites corresponding to all but the N-terminal Met residues of the two monomers in the asymmetric unit. Electron density maps to 2.5 Å resolution were calculated with solvent-flipped SAD phases and initial tracing of the polypeptide chains was carried out with the program ARP/wARP (Perrakis *et al.*, 1999). The initial model was extended by superimposing equivalent structural features of the two independent monomers, and final refinement and model optimization was carried out to 1.8 Å resolution by alternating crystallographic refinement cycles with the program REFMAC (Murshudov *et al.*, 1997) and manual rebuilding with the program COOT (Emsley and Cowtan, 2004). The refined model was subsequently used to solve the structure of the second crystal form using molecular replacement methods (Navaza, 1994), which was refined as described above. The parameters for the final refinement cycles are shown in Table 1.

Acknowledgements

We thank Steve Busby, David Grainger, Douglas Hurd and Marcus Harrison from OGT, Thierry Garnier, Cedric Fiez Vandal, Vincent Bondet, Jacques Bellalou, Nadine Honoré, Marc Monot and Swapna Uplekar for advice, support, help with early experiments and data analysis. This work was funded by the European Commission (LHSP-CT-2005-018923, HEALTH-F3-2007-201762).

References

Anonymous (2006) *The Global Plan to Stop TB 2006–2015.* Geneva: World Health Organization.

Boshoff, H.I., Myers, T.G., Copp, B.R., McNeil, M.R., Wilson, M.A., and Barry, C.E., 3rd (2004) The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *J Biol Chem* **279:** 40174–40184.

Chambers, H.F., Moreau, D., Yajko, D., Miick, C., Wagner, C., Hackbarth, C., *et al.* (1995) Can penicillins and other beta-lactam antibiotics be used to treat tuberculosis? *Antimicrob Agents Chemother* **39:** 2620–2624.

Chambers, H.F., Turner, J., Schecter, G.F., Kawamura, M., and Hopewell, P.C. (2005) Imipenem for treatment of tuberculosis in mice and humans. *Antimicrob Agents Chemother* **49:** 2816–2821.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393:** 537–544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R. *et al.* (2001) Massive gene decay in the leprosy bacillus. *Nature* **409:** 1007–1011.

DeLano, W.L. (2002) *The PyMol Molecular Graphics System.* Palo Alto, CA: DeLano Scientific.

Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126–2132.

Gandhi, N.R., Moll, A., Sturm, A.W., Pawinski, R., Govender, T., Lalloo, U., *et al.* (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet* **368:** 1575–1580.

Garcia-Castellanos, R., Mallorqui-Fernandez, G., Marrero, A., Potempa, J., Coll, M., and Gomis-Ruth, F.X. (2004) On the transcriptional regulation of methicillin resistance: MecI repressor in complex with its operator. *J Biol Chem* **279:** 17888–17896.

Garcia-Castellanos, R., Marrero, A., Mallorqui-Fernandez, G., Potempa, J., Coll, M., and Gomis-Ruth, F.X. (2003) Three-dimensional structure of MecI. Molecular basis for transcriptional regulation of staphylococcal methicillin resistance. *J Biol Chem* **278:** 39897–39905.

Godsey, M.H., Baranova, N.N., Neyfakh, A.A., and Brennan, R.G. (2001) Crystal structure of MtaN, a global multidrug transporter gene activator. *J Biol Chem* **276:** 47178–47184.

Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J., and Busby, S.J. (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci USA* **102:** 17693–17698.

Gregory, P.D., Lewis, R.A., Curnock, S.P., and Dyke, K.G. (1997) Studies of the repressor (BlaI) of beta-lactamase synthesis in *Staphylococcus aureus*. *Mol Microbiol* **24:** 1025–1037.

Hugonnet, J.E., and Blanchard, J.S. (2007) Irreversible inhibition of the *Mycobacterium tuberculosis* beta-lactamase by clavulanate. *Biochemistry* **46:** 11998–12004.

Jarlier, V., and Nikaido, H. (1990) Permeability barrier to hydrophilic solutes in *Mycobacterium chelonei*. *J Bacteriol* **172:** 1418–1423.

Jarlier, V., Gutmann, L., and Nikaido, H. (1991) Interplay of cell wall barrier and beta-lactamase activity determines high resistance to beta-lactam antibiotics in *Mycobacterium chelonae*. *Antimicrob Agents Chemother* **35:** 1937–1939.

Karls, R.K., Guarner, J., McMurray, D.N., Birkness, K.A., and Quinn, F.D. (2006) Examination of *Mycobacterium tuberculosis* sigma factor mutants using low-dose aerosol infection of guinea pigs suggests a role for SigC in pathogenesis. *Microbiology* **152:** 1591–1600.

Lim, D., Poole, K., and Strynadka, N.C. (2002) Crystal structure of the MexR repressor of the mexRAB-oprM multidrug efflux operon of *Pseudomonas aeruginosa*. *J Biol Chem* **277:** 29253–29259.

Maxam, A.M., and Gilbert, W. (1977) A new method for sequencing DNA. *Proc Natl Acad Sci USA* **74:** 560–564.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240–255.

Navaza, J. (1994) AMoRe: an automated program for molecular replacement. *Acta Crystallogr D Biol Crystallogr A* **50:** 157–163.

Perrakis, A., Morris, R., and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6:** 458–463.

Rengarajan, J., Bloom, B.R., and Rubin, E.J. (2005) Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc Natl Acad Sci USA* **102:** 8327–8332.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16:** 944–945.

Safo, M.K., Ko, T.P., Musayev, F.N., Zhao, Q., Wang, A.H., and Archer, G.L. (2006) Structure of the MecI repressor from *Staphylococcus aureus* in complex with the cognate DNA operator of mec. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62:** 320–324.

Safo, M.K., Zhao, Q., Ko, T.P., Musayev, F.N., Robinson, H., Scarsdale, N., *et al.* (2005) Crystal structures of the BlaI repressor from *Staphylococcus aureus* and its complex with DNA: insights into transcriptional regulation of the *bla* and *mec* operons. *J Bacteriol* **187:** 1833–1844.

Sassetti, C.M., and Rubin, E.J. (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci USA* **100:** 12989–12994.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48:** 77–84.

Sharma, V.K., Hackbarth, C.J., Dickinson, T.M., and Archer, G.L. (1998) Interaction of native and mutant MecI repressors with sequences that regulate *mecA*, the gene encoding penicillin binding protein 2a in methicillin-resistant staphylococci. *J Bacteriol* **180:** 2160–2166.

Stinear, T.P., Seemann, T., Harrison, P.F., Jenkin, G.A., Davies, J.K., Johnson, P.D., *et al.* (2008) Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res* **18:** 729–741.

Stinear, T.P., Seemann, T., Pidot, S., Frigui, W., Reysset, G., Garnier, T., *et al.* (2007) Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res* **17:** 192–200.

Sun, R., Converse, P.J., Ko, C., Tyagi, S., Morrison, N.E., and Bishai, W.R. (2004) *Mycobacterium tuberculosis* ECF sigma factor *sigC* is required for lethality in mice and for the conditional expression of a defined gene set. *Mol Microbiol* **52:** 25–38.

Terwilliger, T.C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* **55:** 849–861.

Wang, F., Cassidy, C., and Sacchettini, J.C. (2006) Crystal structure and activity studies of the *Mycobacterium tuberculosis* beta-lactamase reveal its critical role in resistance to beta-lactam antibiotics. *Antimicrob Agents Chemother* **50:** 2762–2771.

Wehenkel, A., Bellinzoni, M., Grana, M., Duran, R., Villarino, A., Fernandez, P., *et al.* (2008) Mycobacterial Ser/Thr protein kinases and phosphatases: physiological roles and

therapeutic potential. *Biochim Biophys Acta* **1784:** 193–202.

Wilke, M.S., Hills, T.L., Zhang, H.Z., Chambers, H.F., and Strynadka, N.C. (2004) Crystal structures of the Apo and penicillin-acylated forms of the BlaR1 beta-lactam sensor of *Staphylococcus aureus*. *J Biol Chem* **279:** 47278–47287.

Wittman, V., Lin, H.C., and Wong, H.C. (1993) Functional domains of the penicillinase repressor of *Bacillus licheniformis*. *J Bacteriol* **175:** 7383–7390.

Yeats, C., Finn, R.D., and Bateman, A. (2002) The PASTA domain: a beta-lactam-binding domain. *Trends Biochem Sci* **27:** 438.

Zhang, H.Z., Hackbarth, C.J., Chansky, K.M., and Chambers, H.F. (2001) A proteolytic transmembrane signaling pathway and resistance to beta-lactams in staphylococci. *Science* **291:** 1962–1965.

Zhu, Y., Englebert, S., Joris, B., Ghuysen, J.M., Kobayashi, T., and Lampen, J.O. (1992) Structure, function, and fate of the BlaR signal transducer involved in induction of beta-lactamase in *Bacillus licheniformis*. *J Bacteriol* **174:** 6171–6178.

Zhu, Y.F., Curran, I.H., Joris, B., Ghuysen, J.M., and Lampen, J.O. (1990) Identification of BlaR, the signal transducer for beta-lactamase production in *Bacillus licheniformis*, as a penicillin-binding protein with strong homology to the OXA-2 beta-lactamase (class D) of *Salmonella typhimurium*. *J Bacteriol* **172:** 1137–1141.

Supporting information

Additional supporting information may be found in the online version of this article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.